2023-1663

NONCONFIDENTIAL VERSION

_____

UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT

_____

STUPP CORPORATION, a Division of Stupp Bros., Inc., IPSCO TUBULARS INC., MAVERICK TUBE CORPORATION

Plaintiffs,

WELSPUN TUBULAR, LLC USA,

Plaintiff-Appellee

v.

UNITED STATES,

Defendant-Appellee,

HYUNDAI STEEL COMPANY,

Defendant,

and

SEAH STEEL CORP.,

Defendant-Appellant.

_____

Appeal from the United States Court of International Trade in
No. 15-cv-00334, Judge Claire Kelly

_____

**NONCONFIDENTIAL BRIEF FOR
DEFENDANT-APPELLEE UNITED STATES**

BRIAN M. BOYNTON
Principal Deputy Assistant Attorney General

PATRICIA M. MCCARTHY
Director

s/Claudia Burke
CLAUDIA BURKE
Deputy Director

s/Robert Ralph Kiepura
ROBERT RALPH KIEPURA
Trial Attorney
Commercial Litigation Branch
Civil Division
U.S. Department of Justice
P.O. Box 480
Ben Franklin Station
Washington, DC 20044
Tel.: (202) 305-4436
Fax: (202) 353-0461
Email: Robert.kiepura@usdoj.gov

November 6, 2023   Attorneys for Defendant

TABLE OF CONTENTS

<u>Confidential Material Omitted</u>

Defendant-Appellant's confidential sales information is designated by brackets and omitted from page 22 line 3; page 47 lines 20 and 22; page 48 lines 1, 2, 3, and 4; and page 49 lines 13 and 14.

A diagram representing defendant-appellant's confidential sales information is designated by brackets and omitted from page 51.

TABLE OF AUTHORITIES

Cases

Statutes

Regulations

Administrative Decisions

STATEMENT OF RELATED CASES

Pursuant to Federal Circuit Rule 47.5, counsel for appellee states that he is unaware of any other appeal in or from this civil action that was previously before this Court or any other appellate court, and is also unaware of any case pending before this Court or any other court or agency that will directly affect or be directly affected by this Court's decision in this appeal.

At the same time, counsel for the United States is aware of one other action before this Court that counsel understands is not "related" within the meaning of the Court's rule, but involves legal issues similar to issues raised in this appeal (in the context of the separate records of those cases). The action is *Marmen Inc., et al., v. United States*, Fed. Cir. No. 23-1877, which also challenges aspects of the Department of Commerce's "differential pricing" methodology employed in antidumping duty cases.

In addition, counsel is aware of several trial court cases involving challenges to Commerce's "differential pricing" methodology. These include:

- *Mid Continent Steel & Wire, Inc. v United States*, Ct. Int'l Trade No. 15-00213
- *NEXTEEL Co., Ltd. v. United States*, Ct. Int'l Trade No. 18-00083
- *HiSteel Co., Ltd v. United States*, Ct. Int'l Trade No. 22-00142
- *Resolute FP Canada Inc. v. United States*, Ct. Int'l Trade No. 23-00095
- *Maquilacero S.A. de C.V. et al. v. United States*, Ct. Int'l Trade No. 23-00091
- *Garg Tube Export LLP v. United States*, Ct. Int'l Trade No. 21-00169
- *Sahamitr Pressure Container Plc. v. United States*, Ct. Int'l Trade No. 23-00077
- *Matra Americas LLC v. United States*, Ct. Int'l Trade No. 21-00632
- *Government of Canada, et al. v. United States*, Ct. Int'l Trade No. 23-00187

**This matter may ultimately encompass 7 cases that are not yet consolidated but should be soon. The individual case numbers are Ct. Int'l Trade Nos. 23-00187, 23-00188, 23-00189, 23-00204, 23-00206, 23-00207, 23-00209

STATEMENT OF THE ISSUE

Whether Commerce's use of the Cohen's *d* test as part of its differential pricing analysis (DPA) is reasonable when the test is applied to reported sales data without requiring that data satisfy the statistical assumptions of normality, equal variances, and sufficient observation size?

STATEMENT OF THE CASE SETTING FORTH RELEVANT FACTS

I.   Nature Of The Case

This appeal concerns the final determination in the less-than-fair-value (LTFV) investigation covering welded line pipe from the Republic of Korea. *Welded Line Pipe From the Republic of Korea*, 80 Fed. Reg. 61,366 (Dep't of Commerce Oct. 13, 2015), Appx216-218, and accompanying Issues and Decision Memorandum (Dep't of Commerce Oct. 5, 2015) (IDM), Appx219-301 (collectively, *Final Determination*). Relevant to this appeal is Commerce's determination on a single issue: the use of Cohen's *d* test as part of Commerce's DPA, which requires that the respondent's sales pass three tests (Cohen's *d*, ratio, and meaningful difference tests) before an alternative comparison methodology may be applied.

II.   The Statutory Framework

The Tariff Act of 1930, as amended (the Act), establishes a remedial regime to combat unfair trade practices. The antidumping provisions of the Act provide

relief for domestic manufacturers by imposing duties upon imports of foreign

products that are sold in the United States at less than fair value. 19 U.S.C. § 1673.

By statute, Commerce must evaluate whether imported products are

sold in the United States at unfairly low prices. 19 U.S.C. § 1673; *see also Union*

Steel v. United States, 713 F.3d 1101, 1103 (Fed. Cir. 2013). If Commerce

concludes that the U.S. sales are "at less than fair value" – meaning that the

products are dumped into the U.S. market, it will direct U.S. Customs and Border

Protection to assess an "antidumping duty." *Id.* A sale is at "less than fair value"

when, in general, the price charged in the U.S. market – the "U.S. price" – is less

than the price charged in the home market – the "normal value." *See* 19 U.S.C. §

1677b(a)(l)(B)(i); 19 U.S.C. §§ 1677a(a), (b). The antidumping duty is equal to

the "amount by which the normal value exceeds the {U.S. price} for the

merchandise." 19 U.S.C. § 1673.

In determining whether subject merchandise is being sold at less than fair

value, Commerce normally compares "the weighted average of the normal values

to the weighted average of the export (and constructed export prices) for

comparable merchandise" unless it determines another method is appropriate. 19

U.S.C. § 1677f-1(d)(1)(A)(i); 19 C.F.R. § 351.414(c)(1). Under this average-to-

average ("A-to-A") method, Commerce compares the weighted average of a

respondent's comparison market sale prices during the investigation period to the

weighted average of the respondent's sales prices in the United States during the same period. *See* 19 C.F.R. § 351.414(b)(1).

One downside of the A-to-A method is that it may fail to detect instances of "targeted" or "masked" dumping, which occurs when an exporter sells at a dumped price to some customers, regions, or time periods, while selling at higher prices to other customers, regions, or time periods. *See Stupp Corp. v. United States*, 5 F.4th 1341, 1345 (Fed. Cir. 2021) (*Stupp III*) (citing *Apex Frozen Foods Priv. Ltd. v. United States*, 862 F.3d 1337, 1341 (Fed. Cir. 2017) (*Apex II*)). When Commerce uses the A-to-A method, higher-priced U.S. sales can mask these lower-priced targeted U.S. sales that are dumped, which could potentially leave the domestic industry without a remedy from unfair trade practices. *See* Uruguay Round Agreements Act, Statement of Administrative Action, H.R. Doc. No. 103-316, vol. 1 (SAA) at 842 ("In part, the reluctance to use an average-to-average methodology has been based on a concern that such a methodology could conceal 'targeted dumping.' In such situations, an exporter may sell at a dumped price to particular customers or regions, while selling at higher prices to other customers or regions."). In an investigation, for example, if the full extent of dumping is masked, the domestic industry may not receive the relief that the statute affords when the calculated weighted-average dumping margin based on the A-to-A method falls below the two percent *de minimis* threshold.

Congress addressed this problem by enacting 19 U.S.C. § 1677f-1(d)(1)(B).

See Apex II, 862 F.3d at 1342. Section 1677f-1(d)(1)(B) allows Commerce to

compare "the weighted average of the normal values to {U.S. prices} of individual

transactions for comparable merchandise if (i) there is a pattern of {U.S. prices}

for comparable merchandise that differ significantly among purchasers, regions or

periods of time, and (ii) {Commerce} explains why such differences cannot be

taken into account using {the A-to-A method or transaction-to-transaction

method}." 19 U.S.C. § 1677f-1(d)(1)(B)(i)–(ii). But Congress did not specify how

Commerce was to determine whether there exists a pattern of prices that differs

significantly among purchasers, regions, or time. The SAA explains that

Commerce should proceed "on a case-by-case basis, because small differences

may be significant for one industry or one type of product, but not for another."

SAA at 843.

To determine whether there is a pattern of prices that differs significantly,

Commerce devised the DPA which it first used in 2013. *See Xanthan Gum from

the People's Republic of China*, 78 Fed. Reg. 33,351, 33,352 (Dep't of Commerce

June 4, 2013) and accompanying IDM at Comment 3; *Xanthan Gum from Austria*,

78 Fed. Reg. 33,354, 33,355 (Dep't of Commerce June 4, 2013).[1]

---

[1] Prior to 2013, Commerce applied several other tests such as the *Nails* test. *See e.g., Apex II*, 862 F.3d at 1328 ("Applying a court-sanctioned methodology

This analysis contains three tests to address the statutory requirements under

19 U.S.C. § 1677f-1(d)(1)(B). First, Commerce applies the "Cohen's *d* test" to

measure whether average prices to a particular purchaser, region, or time period

differ significantly from average prices to all other purchasers, regions or time

periods. *Stupp Corp. v. United States,* 619 F. Supp. 1314, 1322 (Ct. Int'l Trade

2023) (*Stupp IV* (Appx1-27)). Taken from statistical analysis, the Cohen's *d* test

measures effect size—a means of quantifying the size of the difference between

two groups. Second, the ratio test calculates the proportion of the respondent's

U.S. sales, by value, which "pass" the Cohen's *d* test, to determine if a "pattern"

exists. *Id*. If 33% or less of respondent's sales pass, then Commerce finds that no

pattern exists and uses the A-to-A method. If 66% or more of the respondent's

U.S. sales pass, then Commerce finds that a pattern exists and uses the A-to-T

method, subject to the third test—the meaningful difference test. *Id*. If more than

33% but less than 66% of the respondent's sales pass, then Commerce finds that a

pattern exists but takes a hybrid approach, applying the A-to-T method to those

U.S. sales passing the Cohen's *d* test, and the A-to-A method to the remainder, but

only if the meaningful difference test is met. *Id*. Third and finally, if more than 33

---

knows as the Nails test, Commerce identified for Apex a pattern of targeted sales
that differed significantly from prices of non-targeted sales." (internal citations
omitted)).

percent of sales passed the ratio test, Commerce conducts the "meaningful difference" test to whether the A-to-A method can account for the disparate pricing. Appx14-15. Commerce applies the meaningful difference test by comparing a respondent's weighted-average dumping margin using both A-to-A method and the appropriate alternative comparison method. *Id*. If the A-to-A rate is below the *de minimis* threshold and the rate from the alternative comparison method is greater than the *de minimis* threshold, or if both rates are above the *de minimis* threshold and differ by 25% or more, Commerce may then resort to use of the alternative comparison method to calculate the respondent's weighted average dumping margin. Otherwise, Commerce applies the A-to-A method to all U.S. sales. *Id*.

This appeal concerns Commerce's use of the Cohen's *d* test, which involves comparing the product-specific prices of a "test group" to the prices of the "comparison group." *Stupp III*, 5 F.4th at 1346. For each purchaser, region, and time period, Commerce segregates the respondent's U.S. sales for comparable merchandise into two groups, where the "test group" includes all sales of comparable merchandise to a given purchaser, region, or time period, and where the "comparison group" includes all other sales of comparable merchandise. *Id*. For example, Commerce would segregate sales of comparable merchandise to a particular customer (a test group) and compare them to all remaining sales of

comparable merchandise to all other customers (a comparison group). Commerce

then calculates the means and standard deviations of the test and comparison

groups. *Id.* Commerce also calculates a Cohen's *d* coefficient by dividing the

difference in the groups' means by an average of the groups' standard deviation.

Id.

Accordingly, $d = |\text{mean of test group} - \text{mean of control group}| \div \text{standard}$

deviation. Commerce uses a modified version of this formula, substituting for the

"standard deviation" the root mean square of the groups' individual standard

deviations. This Cohen's *d* coefficient represents a measure of "effect size." *See*

generally Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*,

(2nd ed. 1988), A-580-876, PRRD 8, barcode 4181776-01 (Nov. 12, 2021)

(Appx3728). "Effect size quantifies the size of the difference between two groups

and may therefore be said to be a true measure of the significance of the

difference." Coe, Robert, *It's the Effect Size Stupid: What Effect Size Is and Why It*

Is Important, paper presented at the Annual Conference of the British Educational

Research Association at 7 (September 2002), A-580-876, PRRD 8, bar code

4181776-01 (Nov. 12, 2021) (Appx4337).

Commerce calculates the Cohen's *d* coefficient for each test group to

determine whether the prices to that test group differs significantly from all other

prices of comparable merchandise. If the *d* value of a test group is equal to or

greater than the "large threshold," or 0.8, the prices within that test group are found

to have "passed" the Cohen's *d* test and to differ significantly. As explained

earlier, even if the sale prices of a particular test group pass the Cohen's *d* test, it

does not mean that Commerce will apply an alternative comparison method

because the sales passing Cohen's *d* test must be sufficiently numerous to also pass

the ratio test, *i.e.*, that a pattern exists. Furthermore, Commerce must determine

that the standard A-to-A method cannot account for the price differences that are

the result of the respondent's pricing behavior in the U.S. market. This Court has

sustained both the ratio and meaningful difference tests. *Stupp III*, 5 F.4th at 1355-

56.

III. Course Of Proceedings

C.   The LTFV Investigation

In 2014, in response to antidumping petitions by the domestic industry,

Commerce initiated LTFV investigations on welded line pipe from Korea and

Turkey. *See Welded Line Pipe from the Republic of Korea and the Republic of

Turkey*, 79 Fed. Reg. 68, 213 (Dept. of Commerce November 14, 2014) (*Initiation

of LTFV Investigations*). Commerce selected Hyundai HYSCO (HYSCO) and

SeAH Steel Corporation (SeAH) for individual examination as mandatory

respondents. In October 2015, Commerce issued its final affirmative

determination of sales of welded line pipe from Korea at less than fair value. *Final Determination*, Appx216-218.

In its *Final Determination*, Commerce determined that 39.72 percent of SeAH's U.S. sales exhibited a pattern of prices for comparable merchandise that differed significantly among purchasers, regions, or time periods and that A-to-A method could not account for the price differences resulting from SeAH's pricing behavior in the U.S. market. Appx222 & n.11. Accordingly, Commerce used a "mixed alternative method," in which it applied the A-to-A method to the U.S. sales that did not pass the Cohen's *d* test and the A-to-T method to the U.S. sales that did pass the Cohen's *d* test when calculating SeAH's weighted-average dumping margin. Appx222.

Commerce calculated weighted-average dumping margins of 6.19 percent and 2.53 percent for HYSCO and SeAH, respectively. Commerce subsequently amended the *Final Determination* to correct a ministerial error in HYSCO's calculations, which did not affect SeAH's weighted-average dumping margin. Appx197-215. These consolidated actions followed.

D.   Proceedings Before The Trial Court

2.   The Trial Court's First Opinion: *Stupp I*

On January 8, 2019, the trial court sustained the *Final Determination* with respect to Commerce's DPA. *Stupp Corp et al., v. United States*, 359 F. Supp. 3d

1293, 1299-1308 (Ct. Int'l Trade 2019) (*Stupp I*). On March 7, 2019, the trial

court declined SeAH's request to reconsider its decision. *Stupp Corp et al., v.*

United States, 365 F. Supp. 3d 1373 (Ct. Int'l Trade 2019) (*Stupp II*). SeAH

appealed the trial court's final judgment to this Court with respect to Commerce's

rejection of its case brief and with respect to several aspects of the DPA.

5.   This Court's Decision (*Stupp III*)

This Court affirmed the trial court's decision to sustain Commerce's *Final*

Determination with respect to most of SeAH's challenges. *Stupp III*, 5 F.4th at

1344. However, the Court remanded "the issue of whether it was reasonable for

Commerce to apply a portion of its analysis – specifically, the 'Cohen's *d* test' – to

sales data that may have been of insufficient size, not normally distributed, and

lacking roughly equal variances." *Id*. First, this Court was generally concerned

that:

> Commerce's application of the Cohen's *d* test to data that do not
> satisfy the assumptions on which the test is based may undermine the
> usefulness of the interpretive cutoffs. In developing those cutoffs,
> including the 0.8 cutoff, Professor Cohen noted that "we maintain the
> assumption that the populations being compared are normal and with
> equal variability, and conceive them further as equally numerous."

Id. at 1357 (internal citation omitted). The Court's concern was based on certain

academic articles, which were not on the administrative record or before

Commerce for its *Final Determination*. *Id*. at 1348-51 & 1357-58. Second, the

Court found that "test groups consisting of very few observations may be

particularly problematic." *Id.* at 1358. The Court was also concerned when "test groups contain{} sales prices that hover around the same value" and "are not normally distributed." *Id.* at 1359. Finally, the Court questioned whether Dr. Cohen's thresholds can be applied without following the statistical assumption of normality. *Id*. at 1360.

Therefore, this Court remanded "to give Commerce an opportunity to explain whether the limits on the use of the Cohen's *d* test prescribed by Professor Cohen and other authorities were satisfied in this case or whether those limits need not be observed when Commerce uses the Cohen's *d* test in less-than-fair-value adjudications." *Id*. at 1360. In that regard, the Court expressly invited "Commerce to clarify its argument that having the entire universe of data rather than a sample makes it permissible to disregard the otherwise-applicable limitations on the use of the Cohen's *d* test." *Id*.

6.   The Remand Proceeding

In remanding this case, the Court relied on certain academic materials, which were not on administrative record but were discussed in SeAH's appellate briefs. Commerce's regulations preclude Commerce from considering factual information outside of the administrative record, 19 C.F.R. § 351.302(d). For this reason, on remand, Commerce reopened the administrative record, instructing SeAH to supplement it with the academic materials that SeAH referenced in its

appellate briefs, portions of which the Court discussed and referenced in remanding this issue. Appx35. Other interested parties were given an opportunity to rebut, clarify, or correct information that SeAH submitted. *Id.*; *Cf.* 19 C.F.R. § 351.301(c)(v). This resulted in the submission of over 3,000 pages of new factual information.

On April 4, 2022, the remand proceeding culminated in a 74-page remand redetermination, in which Commerce explained why it is reasonable to apply the Cohen's *d* test without following statistical assumptions. Appx28-101. The "statistical criteria do not serve as a basis for Dr. Cohen's thresholds" and that "the academic literature, which the CAFC referenced, address circumstances which are outside of the context in which Commerce utilizes its Cohen's *d* test." Appx37. The Cohen's *d* test is not dependent upon a determination that the data satisfy the three statistical criteria as asserted by SeAH. Appx73.

In general, effect size is a characteristic of a population, and like other population parameters, requires no statistical significance (which is dependent upon the statistical criteria) the way it would if it were estimated based on sampled data. Appx42-43. Further, Dr. Cohen's thresholds were not developed using the statistical criteria cited by SeAH; instead, Dr. Cohen proposed qualitative small, medium and large thresholds, which, while "arbitrary," Dr. Cohen believed that "the proposed convention will be found to be reasonable by reasonable people."

Appx44 (internal citation omitted). Lastly, the Cohen's *d* test is the first step of the DPA, which also includes the ratio test and meaningful difference tests. The alleged potentially anomalous results in the first step of the analysis, in the context of the entirety of the DPA, are inconsequential because the remaining steps of the analysis compensate for the alleged possible inaccuracies. Appx84.

7.   The Trial Court's Second Decision (*Stupp IV*)

On February 24, 2023, the trial court sustained Commerce's *Remand Redetermination*, holding that "Commerce has adequately explained how its methodology is reasonable." Appx26-27. Specifically, the trial court explained that this Court "identified three potential scenarios in which use of Cohen's *d* could be problematic: first, when the distribution of a respondent's sales data is not normal, second, when the test groups have few data points, and third, when there is minimal variance in a respondent's sales." Appx16. The trial court found that "Commerce reasonably explains that Cohen's *d* test does not operate in a vacuum, but as part of the DPA as a whole." Appx17.

Addressing concerns about population size, the trial court agreed with Commerce's reasoning that "even if the Cohen's *d* values of small test groups were less accurate than for large test groups, this difference does not by itself render Commerce's use of Cohen's *d* test unreasonable, because the ratio test and meaningful difference test compensate for inaccuracies." Appx19. The trial court

- 13 -

also found that even though parties devoted a portion of their briefing to the

academic literature concerning the issues of permissibility of applying Cohen's *d*

test to the total population and distinction between practical and statistical

significance, neither question is determinative of whether Commerce's

methodology is reasonable. Appx17-18.

Discussing the concern whether samples without normal distribution could

potentially result in inappropriate number of passes—that the Cohen's *d* test would

indicate that there was a practical difference in sales prices when in fact none

existed—the trial court similarly held that Commerce reasonably explained that

even if the Cohen's *d* test could inaccurately produce positive results under

unusual circumstances, this possibility does not mean its use of Cohen's *d* test is

unreasonable when combined with the ratio test and meaningful difference test.

Appx20. Regarding Commerce's selection of 0.8 threshold identified by Dr.

Cohen as measure of significant price differences, the Court found that

"Commerce considers a significant difference to be grossly perceptible in the same

way that Dr. Cohen identified a large threshold as one that is 'grossly

perceptible.'" Appx22. It also found that "Commerce's choice of a measurement

that is a function of standard deviation as a uniform approach to identify

differences as significant is reasonable, even if the absolute difference in means is

small." *Id*. (citing SAA at 842-843). The trial court further found that Commerce

does not define significance solely as "grossly perceptible" but also as the

difference it has with respect to respondent's weighted-average dumping margins.

Id. Because the Cohen's *d* test is applied with the ratio and meaningful difference

tests, the use of "the 0.8 percent threshold leads to a relatively few determinations

of targeted dumping." Appx23 (citing Appx59). This observation is based on the

combined effect of the use of the conservative 0.8 threshold, followed by the ratio

test and meaningful difference test further winnowing the number of instances

where an alternative methodology is utilized. *Id.* The trial court further held that

"Commerce's reference to Cohen's work does not circumscribe its discretion to

choose the same values in a new context, because that choice is itself reasonable."

Appx24.

Discussing the concern that prices with small variances could produce

inaccurate results in the Cohen's *d* test, the trial court found that "Cohen's test

would need to generate enough 'false positives' to overcome the 33% threshold, at

minimum, and there is no evidence on the record suggesting that price patterns,

such as that proposed by the Court of Appeals, occur with frequency in SeAH's

sales." Appx25. Additionally, the trial court held that Commerce reasonably

explained that in addition to the ratio test, the meaningful difference test would

normally prevent low-variance sales which pass Cohen's *d* test from impacting a

respondent's dumping margins. Appx25-27. This appeal followed.

SUMMARY OF THE ARGUMENT

The question presented to this Court is whether Commerce's use of the Cohen's d test within the context of its DPA as a whole is a reasonable method of implementing the statute. By way of demonstrating its reasonableness, Commerce addressed the three concerns raised by this Court in its prior decision. First, data sets with non-normal distribution do not inevitably lead to inaccurate results, and, even if certain inaccurate results were produced by non-normal data sets, Commerce's DPA as whole reasonably compensates for this possibility. Further, there is no evidence on the record, and SeAH does not argue, that SeAH's sales produced inaccurate results without considering a normal distribution with the sufficient frequency as to change the outcome of Commerce's analysis. Second, and similarly, Commerce addressed the Court's concern that data sets with small numbers of sales could result in inaccuracies by explaining that the ratio test and meaningful difference test would compensate for any potential issues. Third, Commerce addressed this Court's concern and hypothetical discussing data sets with small variance by explaining that the Cohen's d test continues to operate effectively in that circumstance and as intended, consistent with the direction that "Commerce will proceed on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another." SAA at 843. Again, even if there were hypothetical circumstances when data sets with

small variances did produce inaccurate results in the Cohen's *d* test, the ratio test

and meaningful difference test would generally compensate for that theoretical

situation.

Commerce also addressed the Court's concerns related to the continued

effectiveness of Dr. Cohen's thresholds if a data set is not normally distributed by

explaining that the analysis underpinning the 0.8 threshold is not tied to Dr.

Cohen's discussion of appropriate sample selection. Those thresholds are a

measure of practical significance, not statistical significance. Dr. Cohen developed

his thresholds independently of his assumptions regarding normal distribution,

equal variance, and sufficient numbers of data points as part of a statistical or

"power" analysis. This allows Commerce to continue to employ those thresholds

when measuring the practical significance of the difference in two populations.

SeAH's remaining arguments are without merit. Simply because SeAH can

concoct a contrived hypothetical which purports to demonstrate flaws in the

Cohen's *d* test does not make the DPA unreasonable. Similarly, as demonstrated

by historical data, Commerce's DPA predominantly results in the application of

the standard A-to-A comparison methodology. Finally, both SeAH and amici rely

on sources and arguments not on the record and not presented to Commerce.

These sources and arguments should be disregarded.

ARGUMENT

I.   Standard Of Review

In reviewing the trial court's judgments *de novo*, this Court "reappl{ies} the statutory standard of review that the Court of International Trade applied in reviewing the administrative record." *Nippon Steel v. United States*, 337 F.3d 1373, 1379 (Fed. Cir. 2003) (citation omitted.)

This Court's "precedents make clear that the relevant standard for reviewing Commerce's selection of statistical tests and numerical cutoffs is reasonableness, not substantial evidence." *Stupp III*, 5 F.4th at 1353. In situations involving complex technical and methodological choices, Commerce has wide discretion, and the Court only needs to address whether Commerce's methodological choice is reasonable. *Mid Continent*, 940 F.3d at 667 ("In carrying out its statutorily assigned tasks, Commerce has discretion to make reasonable choices within statutory constraints.").

II.   The Trial Court Correctly Sustained Commerce's Remand Redetermination

This Court should affirm the trial court's decision sustaining Commerce's *Remand Redetermination*. While much of what follows delves into academic discussions of statistical analysis, the trial court was correct when it identified the key issue: whether Commerce's methodology is a reasonable means of implementing the relevant statute. Appx20 n10 (*citing Motor Vehicle Mfrs. Ass'n*

of U.S. v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983); *Ceramica Regiomontana, S.A. v. United States*, 636 F. Supp. 961, 966 (Ct. Int'l Tr. 1986), aff'd, 810 F.2d 1137, 1139 (Fed. Cir. 1987)). SeAH does not dispute this point. App. Br. at 15 (*citing Stupp III* at 1353 ("Our precedents make clear that the relevant standard for reviewing Commerce's selection of statistical tests or numerical cut offs is reasonableness, not substantial evidence.")). This Court's prior decision agreed that Commerce's DPA was reasonable in most respects, but remanded for Commerce to address certain, specified concerns. *Stupp III*, 5 F.4ᵗʰ at 1357-59.

This Court identified three scenarios in which use of Commerce's Cohen's *d* test could be potentially problematic: first, when the distribution of a respondent's sale prices is not normal, second, when the test group has few data points, and third, when there are minimal variances in a respondent's sale prices. *Id*. This Court specifically provided Commerce with the opportunity to explain whether the statistical assumptions regarding normality, equal variances, and sufficient observation size "need not be observed when Commerce uses the Cohen's *d* test in less-than-fair-value adjudications." *Id*. at 1360. This is exactly what Commerce did on remand.

F.   The DPA Accounts For Any Potential Inaccuracies

The question before this Court is whether Commerce's use of the Cohen's *d* test as part of the DPA as a whole is a reasonable means of implementing § 1677f-1(d)(1)(B). *See Ceramica*, 636 F. Supp. at 966 ("As long as the agency's methodology and procedures are reasonable means of effectuating the statutory purpose, and there is substantial evidence in the record supporting the not impose its own views as to the sufficiency of the agency's investigation or question the agency's methodology"). We address each of this Court's concerns individually below: that without the use of statistical assumptions, small sample sizes, or small variances without normal distribution could potentially exaggerate "dumping margins" by introducing an upward bias to effect size. *Stupp III*, 5 F.4th at 1357-60. An overarching point of Commerce's detailed responses to these concerns reiterates how the Cohen's *d* test functions as part of the DPA, operating together with the ratio test and meaningful difference test. Appx53, Appx55, Appx57-58, Appx68-69, Appx81-87, Appx91. The trial court agreed, noting that Commerce's "Cohen's *d* test does not operate in a vacuum . . . ." Appx17-18. Instead, the Cohen's *d* is only the first step in Commerce's three-part DPA. *Id.*

Commerce utilizes the entire DPA when evaluating a respondent's pricing behavior to detect masked dumping, *i.e.*, Commerce never intended for the Cohen's *d* test to operate independently. Appx53-55; *see also* Appx19 ("SeAH's

attacks on Cohen's *d* test presuppose that what SeAH claims are 'false positives' automatically affect the accuracy of Commerce's DPA, when in fact Commerce has allowed for 33% positives before there is any potential effect on a respondent's dumping margins."). As Commerce explained, "{t}he fact that U.S. prices differ does not necessarily mean that dumping is being masked or even that the U.S. sale prices are dumped." Appx54. Thus, the fact the prices of a certain group of sales pass the Cohen's *d* test does not, on its own, mean that Commerce will switch to an alternative comparison methodology.

Commerce requires that at least 33 percent of the value of the U.S. sales must pass the Cohen's *d* test to find that a pattern existed under the ratio test, and that the relative change in the weighted-average dumping margins calculated using the A-to-A method and an alternative comparison method must be at least 25 percent or cross the *de minimis* threshold for the alternative comparison method to be applied. Appx88-91. Even if a small number of observations without normal distribution or with small variances could result in an upward bias to the calculated Cohen's *d* coefficient which potentially could satisfy the large threshold, or small values passing the Cohen's *d* test, it would unlikely result in the application of an alternative comparison method because of the additional requirements of the ratio and meaningful difference tests. *Id*. In an antidumping proceeding, U.S. sales data routinely contains thousands of sales and, even if a small number of sales

potentially might incorrectly pass the Cohen's *d* test, it would unlikely change the outcome of the DPA.

SeAH's reported U.S. data contains [numbers] U.S. sales and there is no evidence that the alleged fact patterns exist in a sufficient number of SeAH's U.S. sales to materially affect the outcome of SeAH's differential pricing test. Appx1584. The trial court found that the "Cohen's {*d*} test would need to generate enough 'false positives' to overcome the 33% threshold, at minimum, and there is no evidence on the record suggesting that price patterns, such as that proposed by the Court of Appeals, occur with frequency in SeAH's sales." Appx25. [2]

G.   **Commerce's Cohen's *d* Test Is Reasonable Even With Non-Normally Distributed Data**

As an initial matter, we note that despite the standard of review, SeAH contends throughout its brief that, because the academic literature does not address the circumstances when Commerce's Cohen's *d* test is applied to the entire population of sale prices in the context of DPA, Commerce's use of it in this manner is improper. SeAH Br. at 18. Further, SeAH claims that Commerce's reasoning is "purely negative" because none of the academic literature "explicitly addresses the use of Cohen's *d* when an entire population, and not just a sample, is

---

[2] Even if the "false positives" were to clear this 33 percent threshold, which they do not, they would need to satisfy the separate requirements of the meaningful difference test.

being analyzed" and that Commerce "has not identified any texts that support its proposed use of Cohen's *d.*" *Id.* at 18-19.

SeAH's "purely negative" argument is also a concession—nothing in the academic literature specifically states that it is *improper* to use Cohen's *d* in the manner utilized by Commerce. Thus, as noted above, the only question before the Court is not whether a particular academic text is controlling on this question but whether Commerce's approach was reasonable. *See Ceramica*, 636 F. Supp. at 966. Commerce demonstrated on remand that its approach was reasonable, and none of SeAH's arguments discussing academic literature refute that point. Appx20.

2.   Cohen's *d* Is Reasonable When Combined With The Ratio Test And Meaningful Difference Test

This Court described a general concern whether Cohen's *d* is still a suitable measure of effect size if the data being analyzed is not normally distributed. *Stupp III*, 5 F.4th at 1357-58. Commerce responded, and the trial court correctly agreed, that the lack of a normal distribution does not inherently make it more likely that sale prices will pass the Cohen's *d* test. Appx20-21. Rather, as Commerce observed, the data involving a heavy-tailed distribution in the sample, as in examples discussed in the *Coe* and *Algina* literature, may result in an effect size that *understates* the effect size if the data had been normally distributed, which contradicts SeAH's claim that the absence of a normal distribution inevitably

- 23 -

results false positives. Appx56. In other words, if the sales data had a heavy-tailed distribution, *fewer* sales would pass the Cohen's *d* test than if the data had been normally distributed. *Id.* Consequently, it becomes *less* likely in that scenario that the ratio test and subsequent meaningful difference test would result in an alternative comparison methodology. *Id.* Commerce also acknowledged that a thin-tailed distribution would have the opposite effect. Appx56. However, it is necessary to point out that the concerns highlighted by *Coe* and *Algina*—cited by this Court in its prior opinion—refer to the robustness of Cohen's *d* when a *sample* of data has a non-normal distribution. *Stupp III*, 5 F.4th at 1358. This concern over robustness relates to using Cohen's *d* on a non-normal sample of a normal population. *Id.* In other words, the non-normality of the sample is only a concern if it differs from normality of the entire population. Because Commerce is analyzing the entire population, the concern highlighted by *Coe* and *Alginia* is not relevant.

Additionally, even if the Cohen's *d* analysis were to overstate the effect size under certain scenarios, while understating it in other situations, the inappropriate affirmative passes, net of the inappropriate negative passes, would have to occur more than 33% of the time in order to have any impact on the ratio test. Appx25; Appx26 ("SeAH does not argue that it received an alternative {comparison} method because its own combined sales inappropriately passed Cohen's *d* test.");

see also Wash. State Grange v. Wash. State Republican Party, 552 U.S. 442, 450

(2008) (stating in the context of a facial challenge to a statute "we must be careful

not to go beyond the statute's facial requirements and speculate about

'hypothetical' or 'imaginary' cases."); *Fieger v. Michigan Supreme Court*, 553 F.

3d 955, 961 (6th Cir. 2009).

3.   Commerce Reasonably Employed The Cohen's *d* To Measure Effect Size Using A Full Population

As the first part of its three part test to implement § 1677f-1(d)(1)(B),

Commerce employs the Cohen's *d* test to measure whether the difference in prices

between two full populations of sale prices are significant. Appx14.   SeAH

claims that the Dr. Cohen's *d* coefficient and his thresholds are "solely for the

purpose of measuring the 'power' of an experiment in certain specified

circumstances," and, in general, other measures of effect size "are designed for use

with a specific type of test of statistical significance." *Id.* at 21-22. Contrary to

SeAH's contentions, the academic literature describes effect size and Dr. Cohen's

d coefficient in the context of the full population of data as well as its application

in the context of a statistical analysis with sampled data. Dr. Cohen explicitly

presents effect size in the context of a population: the general definition of effect

size as "the *degree* to which the phenomenon is present in the population"

(Appx3753 (emphasis in original)):

- 25 -

i.e., *the effect size* (ES) *is some specific nonzero value in the population.* The larger this value, the greater the degree to which the phenomenon under study is manifested. Appx3754 (emphasis in original).

Thus, Dr. Cohen contradicts SeAH's argument: effect size is a phenomenon of the population and it exists independent of any statistical analysis based on data sampled from that population. *Id.*

When an analysis is based on sampled data, the statistical criteria are necessary for determining the statistical significance of the estimated parameters and the relevance of the estimated results. Appx39 (quoting Appx4388 ("A statistically significant result is one that is unlikely to be the result of chance.")). However, when the analysis is based on the full population of data, the calculated parameters are the *actual* values and not estimates, and the accuracy of the calculated parameters does not depend on the statistical criteria. Appx39-41.

Given that Commerce's DPA examines the entire population by calculating actual parameters of the population, Appx39-41, Commerce need not consider the statistical criteria to establish that its calculated results are representative of the entire population. As Dr. Ellis explained, researchers seek to "identify samples that are representative of broader populations" and "use inferential statistics to determine whether sample-based observations reflect population-level parameters." Appx4389. This is where the statistical criteria come into play to ensure that samples are representative of the populations, and sample-based calculations (*i.e.*,

estimates) reflect the population's actual parameters. In contrast, Commerce applies Cohen's *d* test to the entire population and calculates the *actual* population parameters, *e.g.*, the Cohen's *d* coefficient.

SeAH also questions whether Commerce's Cohen's *d* test is based on the full population of data, claiming that there are other prices in other periods, or perhaps other products that should be included. SeAH Br. at 17 & fn. 20. However, SeAH fails to recognize that the purpose of the DPA is to examine whether Commerce may use an alternative comparison method to calculate a respondent's weighted-average dumping margin. That margin calculation is based on a specific population of sales for a given respondent, for subject merchandise, during a defined time period (the period of investigation or review). The Cohen's *d* test is limited to this same universe of U.S. sales. The respondent's U.S. sales are grouped into a test group and a comparison group, which are defined and limited by, respectfully, the sale prices of comparable merchandise to a given purchaser, region, or time period, and all sale prices to other purchasers, regions, or time periods.

H.   Data With Small Variance Do Not Produce Inaccurate Results

Second, Commerce addressed this Court's concern that prices with small variances (*i.e.*, near the same value) could artificially inflate the value of the Cohen's *d* for the relevant sales leading to an inappropriate number of sales that

pass the test. Appx56-59, Appx82-85; *see Stupp III*, 5 F.4th at 1359. This Court

illustrated its concern by posing a hypothetical utilizing data with small price

variances. *Id.* In the Court's hypothetical—all respondent's prices hovered around

$100 and passed Cohen's *d* test—Commerce took the most conservative approach

possible and chose the normal value equal to the highest sales price, which

maximized the hypothetical respondent's dumping margin. *Id*. Even in this

extreme hypothetical scenario, the difference in the mean prices failed the

meaningful difference test and resulted in *de minimis* margins under both A-to-A

and A-to-T methods. *Id*. In other words, while we acknowledge the Court's

legitimate concern, the application of DPA to the hypothetical demonstrates that

the meaningful difference test compensates for the specific issues regarding small

variances leading to an "inflated" Cohen's *d* coefficient. Moreover, as a practical

matter, U.S. sales data generally contain thousands of sales. Accordingly, the

hypothetical scenario that this Court posed (*i.e.*, a small number of sales with small

variance) would generally fail the ratio test, which requires at least 33 percent of

total U.S. sales to have significant difference in prices.

Commerce calculates the actual parameters, and if variances in the two

compared groups of prices are small, then the small difference in the means would

be considered more significant than when variances are larger. Appx86. When a

particular product is priced in such a way that price fluctuations are small, even

small price differences could sway purchasers to buy from a different source. SAA

at 843 (explaining that small differences could be significant for certain industries

and products).

Even if a respondent had unusual pricing patterns that impacted the DPA, it

has a remedy: it could ask Commerce to consider whether the DPA as applied to

the respondent-specific data leads to unreasonable results due to unusual pricing

patterns *in that respondent's data*. However, no such issue exists on the record

before this Court and SeAH made no such request. Appx26.

3.   SeAH's Attacks On The Meaningful Difference Test Are Unavailing

SeAH next ignores this Court's prior decisions and again attempts to attack

the meaningful difference test. SeAH Br. at 27. This Court has already affirmed

Commerce's use of the meaningful difference test in its prior decision *in this case*

as well as others. *Stupp III*, 5 F.4th at 1356 (*citing Apex II*, 862 F.3d at 1347).

Nevertheless, SeAH challenges Commerce's use of the test because SeAH believes

that its use treats as meaningful, small price differences. The Court should decline

to consider this because it is law of the case at this point.

Even if the Court chooses to consider it, SeAH identifies no actual harm

from the method. SeAH does not dispute that, as a general matter, small variances

in prices of a product would not result in meaningful difference in margins

calculated under A-to-A and A-to-T methods. SeAH Br. at 27. SeAH

acknowledges that, when Commerce specifically addressed this Court's hypothetical involving sales with prices that hovered around the same value, such sales failed the meaningful difference test and resulted in *de minimis* margin under both A-to-A and A-to-T methods. Appx58-59; SeAH Br. at 26-27. Instead, SeAH contends that it is possible to construct another hypothetical scenario—different from that posed by this Court in *Stupp III*—involving multiple products under which the sales with small variances could pass a meaningful difference test. SeAH Br. at 26-27.

SeAH's new hypothetical is contrived and unrealistic: it involves four sales of product 1 to customer A at a weighted-average price of $100 with no price variances within the group and four sales to Customer B at a weighted-average price of $99.99 with 1 cent variance within the group.[3] *Id*. at 28. Of course, if the DPA were to be applied to sales of this product, the outcome would be the same as in the Court's hypothetical which involved *small* price variances in both the test and comparison groups. The small difference in weighted-average prices results in failure of the meaningful difference test and the application of the standard A-to-A method.

---

[3] SeAH's hypothetical assumes that the exporter only has two customers.

SeAH's hypothetical introduces the prices of a second product with *large* price variances to each customer. *Id*. at 29. It assumes that two sales of a second product to each of the same two customers at $90 and $110 with $20 variance within both the test and comparison group. *Id*. Finally, SeAH selects a normal value of $101.97 for each of the two products and argues that it is possible for the combined sales of both products to pass the meaningful difference test. *Id*. at 30.

SeAH contends that its analysis demonstrates that the meaningful difference test does not compensate for all situations involving small variances, when a second product with large variances is added. *Id*. However, SeAH's hypothetical disintegrates when even a single element of its cherry-picked parameters is slightly modified. As the tables below illustrate, if just one additional sale of the second product is made to each customer at the average price of that product such that the prices for the second product continue to not pass the Cohen's *d* test, SeAH's "analysis" fails. Specifically, SeAH's hypothetical will result in a *de minimis* rate regardless of the comparison method:

Cohen's *d* test - product #1

Product #1	Customer 1	Customer 2	Total
Sale #1	99.999	100.000	
Sale #2	100.000	100.000	
Sale #3	100.000	100.000	
Sale #4	100.000	100.000	
Total	399.999	400.000	
Average	99.99975	100.000	
Diff of Means			0.00025
Std Dev	0.000433013	0.000000000	0.000306186
Cohen's *d*			0.816496581

Note: Because there are only two customers, both customers pass.

Cohen's *d* test - product #2

Product #2	Customer 1	Customer 2	Total
Sale #1	90.000	90.000	
Sale #2	110.000	110.000	
Sale #3	100.000	100.000	
Total	300.000	300.000	
Average	75.000	75.000	
Diff of Means			0.000000
Std Dev	8.164966	8.164966	8.164966
Cohen's *d*			0.000000

Note: Because there are only two customers, both customers do not pass.

Ratio Test

Cohen's *d* Test	Sales Value (US$)
Passing	799.999
Not Passing	600.000
Total	1,399.999
Ratio	57.14%

Here, because only 57.14% passes the Cohen's *d* test, the potential

alternative comparison methodology is the "mixed" method. However, Commerce

would still have to employ the meaningful difference test:

Meaningful Difference Test
Normal Value: 101.9700

Product	Sale	Quantity (units) (A)	Value (US$) (B)	Normal Value (US$/unit) (C)	A-to-A Method U.S. Price (US$/unit) (D)=(B)/(A)	Dumping Margin (US$/unit) (E)=(C)-(D)	Weighted Average Dumping Margin (F)=(E)/(B)	Alternative Comparison Method Product #1: A-to-T Product #2: A-to-A U.S. Price (US$/unit) (G)=(B)/(A)	Dumping Margin (US$/unit) (I)=(C)-(G)	Weighted Average Dumping Margin (J)=(I)/(B)
Product #1	Sale #1	1	99.999	101.970	99.9999	1.9701		99.9990	1.9710	
	Sale #2	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #3	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #4	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #5	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #6	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #7	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Sale #8	1	100.000	101.970	99.9999	1.9701		100.0000	1.9700	
	Product Tot	8	799.999							
								A-to-T Total	15.7610	
Product #2	Sale #1	1	90.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Sale #2	1	110.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Sale #3	1	90.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Sale #4	1	110.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Sale #5	1	100.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Sale #6	1	100.000	101.970	100.0000	1.9700		100.0000	1.9700	
	Product Tot	6	600.000							
								A-to-A Total	11.8200	
Total		28	1,399.999			27.5810	**1.9701%**		27.5810	**1.9701%**

NOTE: U.S. Price, under the A-to-A method, is equal to the product-specific total value divided by the product-specific total quantity.

The table above demonstrates that, if SeAH's contrived hypothetical is modified by even one sale of the second product to each customer—a sale at the mean price for the product—the A-to-A method would be applied.

4.   SeAH's Claims Regarding The Ratio Test Are Incorrect

The statute contains a requirement that Commerce determine whether there is "a pattern of export prices" "differ{ing} significantly among purchasers, regions, or periods of time" before selecting the A-to-T method. SeAH believes Commerce (and the trial court) incorrectly concluded that the pattern requirement is "only and exclusively" addressed by the ratio test within the DPA. SeAH Br. at 34-35.

SeAH's argument again ignores the fact that this Court already sustained the ratio test as reasonable approach to establish existence of a pattern. *Stupp III*, 5 F.4th at 1346. This Court explicitly held that "Commerce's ratio test reasonably implements the statutory requirement that Commerce determine whether there is 'a pattern of export prices' 'differ{ing} significantly among purchasers, regions, or periods of time' before selecting average-to-transaction method." *Id.*

Nevertheless, both SeAH and Amici contend that if results of the Cohen's *d* test contain any degree of inaccuracy, they are meaningless, and the results of the ratio test will be inaccurate. SeAH Br.at 33-34; Amici Br. at 32-35. SeAH's argument misrepresents how Commerce's DPA operates. SeAH alleges that neither the ratio test nor the DPA as a whole can distinguish between true patterns and chance fluctuations. SeAH Br. at 36. This argument proceeds from a false premise: Commerce does not examine randomly fluctuating data; rather it examines actual prices that are based on deliberate pricing decisions of an exporter and reflect the exporter's pricing behavior. Appx68-77. The allegation that a company's prices may be set by random chance has no basis in law or fact.

As to the remainder of SeAH's allegation concerning the ratio test, its argument is conclusory. SeAH assumes that the Cohen's *d* test is flawed and because of that flaw, it assumes the ratio test must also be flawed. SeAH at 36. First, as discussed above, the Cohen's *d* test is not flawed. Appx52, Appx43,

Appx86. Second, the ratio test is specifically designed to ensure that, even if some

prices could improperly pass Cohen's *d* test in unusual circumstances, they must

occur more than a third of the time to even have a chance of impacting the DPA as

a whole. Appx19-20 ("SeAH's attacks on Cohen's *d* test presuppose that what

SeAH claims are 'false positives' automatically affect the accuracy of Commerce's

DPA, when in fact Commerce has allowed for 33% positives before there is any

potential effect on a respondent's dumping margins."); Appx88-91. As noted

previously, SeAH does not allege that its own data on the record produced such a

result and is relying on nothing more than conjecture and contrived hypotheticals.

Appx26.

I.   **Commerce's Calculation Of The Actual Parameters Of The Entire Population Addresses Concerns Related To Small Data Sets**

Third, this Court was concerned that small data sets may result in a bias in

the results of the Cohen's *d* test and that such a bias would lead to more passing

results, thereby exaggerating dumping margins. *Stupp III*, 5 F.4th at 1359.

Commerce addressed this concern, explaining that Cohen's *d* test, when applied to

the entire population, is accurate and does not result in a positive or negative bias.

Appx41-43. The calculated *actual* parameters of the population are

mathematically accurate regardless of whether such population has two

observations or two thousand observations. *Id.* Commerce's Cohen's *d* test "*does*

not estimate the Cohen's *d* coefficient, let alone overestimate it, but rather

calculates the actual Cohen's *d* coefficient based on the entire population of sale

prices." Appx52 (emphasis in the original). This is significant because "for the

Cohen's *d* test applied in the context of the {DPA}, there is no estimation of the

parameters (*i.e.*, mean, standard deviation, and effect size) of the test or

comparison group as the calculation of these parameters is based on the complete

universe of sale prices to the test and comparison groups." Appx42-43.

Accordingly, Commerce calculates the actual, mathematically accurate parameters

(mean, standard deviation, and effect size) based on the *entire* population of

relevant observations.

Contrastingly, "statistical criteria observed in academic literature (such as

the number of observations, a normal distribution and approximately equal

variances) are related to the statistical significance of sampled data and establish

the reliability of an *estimated* parameter (*e.g.*, mean) based on the sample data to

be within a determined confidence interval of the actual parameter." Appx42

(emphasis added). Put differently, the statistical assumptions ensure that a sample

is sufficiently representative of the entire population, which is not a concern when

the entire population is considered. Commerce explained, "each time these

parameters would be calculated as part of Commerce's Cohen's *d* test, the exact

same results would be found because the calculated parameters are the parameters

of the entire population and not an estimate of the parameters based on a sample."

Appx43. "{T}here is no bias, positive or negative, in the results of Commerce's

Cohen's *d* test." Appx52, Appx43.

J.   Commerce Reasonably Addressed Concerns About Dr. Cohen's Thresholds

This Court also questioned Commerce's use of the 0.8 threshold identified

by Dr. Cohen when it evaluates data that does not meet the statistical criteria "that

the populations being compared are normal and with equal variability, and

conceive them further as equally numerous." *Stupp III*, 5 F.4th at 1357 (quoting

Appx3764, citing Appx3768-3769, Appx4316 Appx4315-4316). On remand,

Commerce explained that the quotation from Dr. Cohen referred to by this Court

concerned a different calculation meant to provide context for and illustrate his

thresholds. Appx-46-47. Specifically, when Dr. Cohen stated that the "percent

non-overlap" of two curves could only be measured if the requirements of

normality, equal variances, and equally numerous so that the percent non-overlap,

the first of the three "U measures" can be calculated.   *Id.* As discussed below, the

assumptions of normality, equal variances, and numerosity were not used by Dr.

Cohen to define his proposed convention of small, medium, and large thresholds,

but are used to present one approach, i.e., the "U measures," to illustrate, what

each of Dr. Cohen's proposed thresholds mean in terms that one can easily grasp.

Id.; Appx4425 ("Cohen's effect size classes have two selling points. *First, they are*

easy to grasp. You just compare your numbers with his thresholds to get a ready-made interpretation of your result. Second, although they are arbitrary, they are sufficiently grounded in logic for Cohen to hope that his cut-offs 'will be found to be reasonable by reasonable people'" (internal citations omitted, emphasis added))

Dr. Cohen proposed these thresholds as a method of comparing two sets of data, but he acknowledged that the thresholds were "arbitrary" and "qualitative" and that they "run the risk of being misunderstood." Appx3756. But Dr. Cohen believed that "{a}lthough arbitrary, the proposed conventions will be found to be reasonable by reasonable people." Appx3757. Academic literature does not state that the small, medium, and large thresholds resulted from complex statistical models or tests that are based on the statistical criteria. Rather they are "arbitrary" values which Dr. Cohen believed "to be reasonable by reasonable people," and for which Dr. Cohen provided real world examples to illustrate each threshold. Appx44-45 (quoting Appx3757). However, despite that arbitrariness, Dr. Cohen and other academics believed these thresholds to be an effective tool in measuring effect size. *Id.*; Appx4388.

"{T}he purpose of the Cohen's *d* test is to determine the significance of the difference in the prices between a given purchaser, region, or time period and all other sales of the comparable merchandise." Appx44. Commerce uses Cohen's *d* test to measure the *practical* significance of price differences, which is distinct

from *statistical* significance. Appx39-40 (quoting Appx4388 ("It is quite possible,

and unfortunately quite common, for a result to be statistically significant and

trivial. It is also possible for a result to be statistically nonsignificant and

important.")). The concerns regarding normality, variance, and numerosity pertain

to *statistical* significance and are not relevant to question of the *practical*

difference in prices between two sets of data. *Id.*

Commerce employs the 0.8 threshold to identify where prices "differ

significantly" under 19 U.S.C. § 1677f-1(d)(1)(B)(i). Appx38, Appx44-45. As the

trial court noted, the 0.8 threshold represents a difference that is "grossly

perceptible." Appx22, Appx55-56, Appx79. The SAA instructs Commerce to

proceed on a case-by-case basis because small differences could be significant for

one industry or type of product, but not for another. SAA at 843. As the trial court

found, "Commerce's choice of a measurement that is a function of standard

deviation as a uniform approach to identify differences as significant is reasonable,

even if the absolute difference in means is small" and, thus, it is reasonable for

Commerce to tailor its approach to pricing parameters of products. Appx22 (citing

SAA at 842-843).

Commerce also selected the 0.8 threshold as a "conservative" threshold that

it expected to result in a limited application of the alternative comparison

methodology. *Id*. Commerce's choice is supported by the historical data, which

demonstrates that in most investigations when the Cohen's *d* test was applied as part of the DPA, Commerce used the standard methodology. Appx58-59 (finding that the alternative methodology was applied to a relatively small number of respondents).[4] As the trial court noted, Commerce relied upon its Congressionally recognized expertise in the determining when prices are significant. Appx23-24 (*Citing* 19 U.S.C. § 1677f-1(d)(1)(B)(i)). This methodological choice is entitled to judicial deference. *Thai Pineapple Pub. Co. v. United States*, 187 F.3d 1362, 1365 (Fed. Cir. 1999); *Cf. Fujitsu General Ltd. v. United States*, 88 F.3d 1034, 1039 (Fed. Cir. 1996).

---

[4] In some determinations, the application of this conservative threshold resulted in 0.00 percent of the respondent's U.S. sales passing the Cohen's *d* test. *See e.g.*, *Certain Activated Carbon from the Peoples Republic of China*, 83 Fed. Reg. 23,254 (May 18, 2018) (prelim. admin. review) and accompanying PDM at 4(g) (finding that 0.00 percent of Datong Juquiang's U.S. sales passed Cohen's *d* test); *Fresh Garlic from the Peoples Republic of China*, 82 Fed. Reg. 57,718 (Dec. 7, 2017) (prelim. admin. review) and accompanying PDM at "Comparisons to Normal Value" section (finding that for Yudi and Join, 0.00 percent of U.S. sales passed the Cohen's *d* test); *Certain Small Diameter Carbon and Alloy Seamless Standard, Line and Pressure pipe from Romania*, 82 Fed. Reg. 26,452 (June 7, 2017) (prelim. Admin. Review) and accompanying PDM at "A. Determination of Comparison Method" section (finding that for Silcotub 0.00 percent of the value of U.S. sales passed Cohen's *d* test); *Prestressed Concrete Wire Strand from Thailand*, 82 Fed. Reg. 9,197 (Feb. 3, 2017) (prelim. Admin. Review) and accompanying PDM at "B. Results of Differential Pricing Analysis" section (finding that for SIW 0.00 percent of U.S. sales passed Cohen's *d* test).

SeAH—and amici—contend that Commerce failed to "justify its use of Dr. Cohen's proposed thresholds as a matter of statistical practice or as a matter of mathematical logic." SeAH Br. at 43-47; *cf.* Amici Br. at 19-21 (arguing that Dr. Cohen referred to populations rather than samples in describing non-overlap (*U*)). These arguments are incorrect.

At the outset, we note that this Court has already held that Commerce's analysis concerning these thresholds is reasonable. *See Mid Continent*, 940 F.3d at 673 (internal citations omitted) ("The Trade Court described Commerce's rationale for adhering to the 0.8 line and explained why that rationale is reasonable . . . . We agree with the Trade Court that this rationale adequately supports Commerce's exercise of the wide discretion left to it under 19 U.S.C. § 1677f-1(d)(1)(B)."). Dr. Cohen presented the concept of "effect size" that "*is some specific nonzero value in the population.* The larger this value, the greater the degree to which the phenomenon understudy is manifested." Appx3754 (emphasis in original). As quoted by SeAH:

> From one point of view, a universal ES index, applicable to all the various research issues and statistical models used in their appraisal, would be the ideal. Apart from some formidable mathematical-statistical problems in the way, even if such an ideal could be achieved, the result would express ES in terms so unfamiliar to the researcher in behavioral science as to be self-defeating.

SeAH Br. at 45-46, and fn. 67 (quoting Appx6109). Here, SeAH misappropriates

Dr. Cohen's "self-defeating" task by ascribing it to "Professor Cohen's rule-of-

thumb that a d of 0.8 indicates a 'large' effect {that} can be used as a universal

yardstick." SeAH Br. at 43. SeAH misleadingly portrays the use of Dr. Cohen's

large, 0.8, threshold "universally" as counter to Dr. Cohen's intentions when Dr.

Cohen expresses no such limitation. To the contrary, Dr. Cohen established

thresholds for evaluating the magnitude of the effect size which are "easy to grasp"

and "are sufficiently grounded in logic for Cohen to hope that his cut-offs 'will be

found to be reasonable by reasonable people.'" Appx44. (quoting Appx4416).

Instead, what Dr. Cohen described as a "self-defeating" effort would be to

establish a universal measure of effect size, rather than having different measures

of effect size for different applications. Dr. Cohen actually discusses an

application-specific measure of effect size, *e.g.*, the *d* coefficient for the difference

of the means, to avoid the "self-defeating" task of a universal measure of effect

size. In other words, devising a single measure of effect size for all types of

applications would be self-defeating, so Dr. Cohen presents different formulations

of effect size for different applications. Appx3757 ("Each of the Chapters 2-10

will present in some detail the ES index appropriate to the test to which the chapter

is devoted."); Appx4390 ("Effect sizes come in many shapes and sizes. By one

reckoning there are more than seventy varieties of effect size"). Thus, SeAH's

- 42 -

argument, that Dr. Cohen had rejected the general application of his thresholds without some undefined limitations, is a misrepresentation of Dr. Cohen's text.

The effect size is a unitless index which must be interpreted to provide meaning to the user. Appx44 (quoting Appx4416 ("To assess the practical significance of a result it is not enough that we know the size of an effect. Effect magnitudes must be interpreted to extract meaning.")). Dr. Cohen proposed "{f}or each statistical test's ES index, …. *as a convention*, ES values to serve as operational definitions of the qualitative adjectives 'small,' 'medium,' and 'large.'" Appx3756 (emphasis in the original). Dr. Cohen recognized the dangers of such qualitative categories, such as the definitions being arbitrary and misunderstood, although he believed that "the proposed conventions will be found to be reasonable by reasonable people." Appx3756-3757. For the *d* coefficient, Dr. Cohen defined the small, medium and large thresholds as effect size indexes of 0.2, 0.5 and 0.8, respectively. Dr. Cohen also provided examples of real-world comparisons which illustrated each of these index levels. For the "large" threshold, Dr. Cohen identified such comparisons as: the mean IQ difference estimated between holders of the Ph.D. degree and typical college freshmen, or between college graduates and persons with only a 50-50 chance of passing in an academic high school curriculum. Appx3770. "These seem like grossly perceptible and therefore large differences, as does the mean difference in height between 13- and 18-year-old

girls, which is of the same size (d = .8)." Appx3770. Thus, Dr. Cohen's

thresholds were not derived based on quantitative analysis of numerous

comparisons, but were proposed as reasonable index levels which were then

illustrated with study results which presented the proposed index levels as well as

the various "U measures" which are dependent on the statistical criteria to permit

the calculate these three measures. These illustrative examples contribute to or

confirm for reasonable people that Dr. Cohen's proposed conventions are

reasonable, notwithstanding their known shortcomings. *See Mid Continent III*, 940

F.3d at 673 ("The 0.8 standard is "widely adopted" as part of a "commonly used

measure" of the difference relative to such overall price dispersion…").

The discussion of the statistical criteria (*i.e.*, assumptions) in *Cohen* is

unrelated to Dr. Cohen's proposed small, medium, and large thresholds:

> If we maintain the assumption that the populations being compared are
> normal and with equal variability, and conceive them further as equally
> numerous, it is possible to define measures of nonoverlap (U) associated
> with d which are intuitively compelling and meaningful.

Appx3764; Appx46-47; Amici Br. at 11. This text relates to the calculation of the

"U measures," including the percent non-overlap, and the assumptions of

normality and equal variances are required to calculate the area under the two bell

curves which are not common to both distributions. Appx46-47; *see also* SeAH

Br. at 23 (recognizing the unique mathematical properties of a normal distribution

which allow for these non-overlap calculations). The percent non-overlap,

however, *does not define Dr. Cohen's thresholds*; rather, for a given effect size, one can calculate (or look up in *Cohen* Table 2.2.1 under U₁) the equivalent percent non-overlap. For a large, 0.8, effect, for example, it is an interesting factoid that the percent non-overlap is 47.4%. Appx3764, Table 2.2.1, and Appx3769. Similarly, the reference to percentile standing in *Grissom* requires normality and equal variances to calculate this value. Appx47-50 (discussing same issue with regards to *Coe*). Indeed, this is also included as part of Dr. Cohen's percent non-overlap as "U measure" U₃ in Table 2.2.1. Appx3763-3768.

As with the percent non-overlap, the percentile standing *does not define Dr. Cohen's thresholds*; rather, effect size is needed to define the percentile standing. Appx47-50. Accordingly, the assumptions of normality, equal variances, and equally numerous to calculate these various theoretical quantitative measures of non-overlap *are not part* of Dr. Cohen's derivation of his qualitative convention of small, medium, and large thresholds, and such statistical criteria do not limit Commerce use of Dr. Cohen's large, 0.8, threshold in the Cohen's *d* test. *Id.*

As discussed above, Dr. Cohen, in proposing his qualitative thresholds, believed that they "will be found to be reasonable by reasonable people." Appx3756-3757. In addition to the measures of percent non-overlap, Dr. Cohen presented real-world observations to demonstrate that his proposed conventions are reasonable. Dr. Cohen described "grossly perceptible" situations such as the

difference in IQ between Ph.D degree holders and incoming college freshmen or

the difference in the heights of 13- and 18-year old girls. *Id.* For each threshold,

Dr. Cohen provided a qualitative perspective and examples of analysis which

represent each qualitative threshold. For the large threshold—0.8—Dr. Cohen

described this level of difference as "grossly perceptible." Appx3770. Here,

Commerce has applied these same thresholds for the reasons articulated by Dr.

Cohen—these grossly perceptible differences are a reasonable cutoff.

SeAH claims that because IQ scores and girls' heights are normally

distributed, that any situation which uses Dr. Cohen's thresholds must also be

normally distributed. SeAH Br. at 47-49. SeAH logic fails for the same reason as

SeAH linking the statistical assumptions required to calculate the measures of

percent non-overlap with Dr. Cohen's proposed thresholds. Dr. Cohen's examples

of real-world situations falling in the small, medium, and large effect size

categories neither define those qualitative thresholds nor quantify the 0.2, 0.5 or

0.8 values which Dr. Cohen realized were arbitrarily selected. Thus, SeAH's

purported linkage between Dr. Cohen's effect size thresholds and the statistical

criteria is unsupported.

Finally, SeAH attempts to demean the use of Dr. Cohen's thresholds by

arguing that "his proposed thresholds were intended 'for use only when no better

basis for estimating the {effect size} index is available.'" SeAH Br. at 46 (quoting

Appx3768). Nonetheless, that is exactly how this Court affirmed Commerce's use

of Dr. Cohen's thresholds:

> The 0.8 standard is "widely adopted" as part of a "commonly used measure" of the difference relative to such overall price dispersion; and it is reasonable to adopt that measure where there is no better, objective measure of effect size. We agree with the Trade Court that this rationale adequately supports Commerce's exercise of the wide discretion left to it under 19 U.S.C. § 1677f-1(d)(1)(B).

Mid Continent III, 940 F.3d at 673 (internal citations omitted). Therefore, Dr.

Cohen's thresholds are not dependent on the distributions or variances of the data

which are under examination to determine the magnitude of the difference in the

means and are a reasonable tool for interpreting the magnitude of effect size. *See*

Appx4424 ("Cohen's cut-offs provide a good basis for interpreting effect size and

for resolving disputes about the importance of one's results.").

F.   SeAH's Analysis Of Its Sales Data Is Misleading

SeAH next contends the price differences for SeAH's sales are not "grossly

perceptible" when the price data are examined. SeAH Br. at 49-50. Specifically,

SeAH asserts that the sale prices of product (CONNUM) 1-03-03-06-1 to customer

102020 pass the Cohen's *d* test. SeAH Br. at 50 (includes SeAH's business

proprietary sales data for this CONNUM). SeAH contends that it has [ ] sales of
numbers

CONNUM (the product control number) 1-03-03-06-1 to customer 102020, with

an average price of $[numbers], and a standard deviation of [numbers]. *Id.* at 15-

16. The comparison group consisting of U.S. sales of that product to all other

customers included [numbers] transactions, with an average price of $[numbers], and a standard deviation of [numbers]. *Id* at 16. The margin calculation program calculated a difference in average prices of [numbers], a pooled standard deviation of detail below [numbers], and a Cohen's *d* coefficient of 1.00739. *Id.* Assuming that the statistical criteria must be followed, SeAH provides a graph of two bell curves which represents the prices to the test and comparison groups, a "visual examination {of which} appears to show a 'pattern' of price differences" where the prices to the test group "appear{} to be markedly *lower*." SeAH Br. at 51 (emphasis in original). SeAH then provides a scatter diagram that plots the individual prices in the test and comparison groups by date of sale and states that the diagram demonstrates that there is no "clear pattern." SeAH Br. at 52.

SeAH analysis embodies numerous flaws. First, SeAH continues its mischaracterization of the Cohen's *d* test as determining whether there is a "pattern," whereas the test determines whether the price difference is significant. SeAH Br. at 52. Additionally, SeAH attributes the cause of the purported incorrect results to the "artifact of the assumption that prices are Normally {sic} distributed and have equal variances." SeAH Br. at 52. As explained above, Commerce has not made, and need not account for, any such assumption.

Furthermore, SeAH incorrectly equates the "grossly perceptible" qualitative description of a large effect to a "visually discernable" standard. Many types of

information cannot even be visually observed. SeAH Br. at 49-50. For example, for IQ scores of individuals or the prices of welded pipe, one cannot visually see such data.

SeAH's charts do not accurately represent SeAH's data and how the Cohen's *d* test operates. The first chart represents a *hypothetical* normal "bell curve" distribution of prices, which, according to SeAH, shows price differences. *Id*. at 51. Although Dr. Cohen uses a similar graphical representation for sampled data in his discussion of non-overlap, SeAH's graph does not accurately represent SeAH's prices for (CONNUM) 1-03-03-06-1. First, the number of observations in each group, which are plotted on the y-axis of the graph, are not the same in SeAH's narrative description of the price data. Further, the standard deviations in the graph for each group also appear to be identical, whereas the standard deviations for the test and comparison groups in SeAH's example, [numbers] and [numbers,] respectively, are not the same. The only aspect of SeAH's graph that is accurate is that the mean prices for each group appear to be at the correct points on the x-axis. Accordingly, the chart fails to accurately present a usable visualization of prices for control number 1-03-03-06-1. Nonetheless, SeAH concedes that the prices for customer 102020 and CONNUM 1-03-03-06-1 are "markedly *lower*" than prices in the comparison group (all other customers), *i.e.*, they differ significantly. SeAH Br. at 51.

SeAH's second chart (scatter diagram) shows prices of individual sales to each customer group spread across various dates. *Id*. at 52. However, that is not what the measure of effect size examines. The Cohen's *d* test, as a measure of effect size, considers the mean of each group as the representative of each group. Further, the date of individual sales is not relevant to the Cohen's *d* test by purchaser. Commerce calculates the average (mean) price for the relevant period for a given purchaser and for all other purchasers. Accordingly, the frequency of sales, how they are "scattered," or how prices of individual transactions compare to each other on any particular date are not relevant to Commerce's analysis by purchaser. The dates of sales are only relevant to the extent that Commerce includes all sale prices for the relevant time period, *i.e.*, the entire period of investigation; within the relevant time period, it does not segregate U.S. prices by date. The Cohen's *d* test considers differences by purchaser and differences by time period as separate analyses.

In applying the Cohen's *d* test, for each of the two groups (here defined by purchaser), Commerce calculated the *actual* mean (average) of such prices *for the period of investigation* based on the entire population of relevant observations. Commerce compared these two actual means, and the comparison result, relative to the dispersion of prices within the two groups, has exceeded Dr. Cohen's threshold for the large effect size that represents differences that are "grossly perceptible."

Appx3769-3770. This is sufficient for demonstrating that this customer and

CONNUM the average price for the period of investigation differs significantly

from average price of comparable merchandise to all other customers. To illustrate

this point, using the same scale as SeAH did in its chart (SeAH Br. at 52), the

following chart shows that the difference between the average prices to customer

102020 and all other customers is "grossly perceptible:"

<div align="center">

Comparison of SeAH's Period of Investigation
Average (Mean) Price to Customer 102020 and SeAH's
Average Price to the Comparison Group (All Other Customers)

</div>

[

Diagram

].

 To the extent that SeAH contends that on specific dates the relation between

individual transaction prices among these two groups moved in a different

direction, it is not surprising because the prices of individual transactions on a

specific date could fluctuate differently from the overall pricing trend as

represented by the mean price of each group over the entire period of investigation.

However, while Commerce could have decided to compare individual transaction

prices on specific dates while ignoring other prices,[5] it did not adopt the

methodological approach represented in SeAH's charts. Rather, Commerce

selected a methodology that calculates the actual *average* price for the test and

comparison groups as a better measure of the overall pricing behavior.

Accordingly, SeAH's second chart has no bearing on Commerce's Cohen's *d*

analysis.

G.   SeAH Arguments Regarding The Historical Data Concerning The Application Of Differential Pricing Are Meritless

SeAH makes a sweeping claim that Commerce's methodology inherently

results in "false positives"—sales that pass the Cohen's *d* test even though they

have small differences and thus, allegedly, should not pass—and offers

hypothetical examples to demonstrate that the methodology systematically leads to

the use of an "alternative comparison methodology" to "inflate the dumping

margin." *See e.g.,* SeAH Br. at 10, 26, 31-32; *cf.* Amici Br. at 29-30.

---

[5] Comparing the prices of individual transactions on a specific date, instead of comparing the period of investigation averages, would not automatically make the finding of significant difference more or less likely.

However, the data demonstrates that after applying its DPA, Commerce

applied the standard comparison method in most investigations. Appx59.

Commerce analyzed the overall results of its DPA for investigations in the years

2015 and 2021 and found that "Commerce's actual application of the Cohen's *d*

test in the context of the DPA resulted in the application of an alternative

comparison methodology to a relatively small number of respondents." Appx59.

Specifically, in 2015, when this investigation was completed, Commerce applied

its DPA to 18 respondents, which resulted in the application of the alternative

calculation methodology to only four companies, including SeAH. *Id*. In other

words, despite SeAH's assertions that Commerce's methodology is so

undemanding that it creates "false positives," leading to the unwarranted

application of the alternative comparison methodology, the data demonstrates that

in 2015 "only 22 percent of respondents with calculated rates had their weighted-

average margin calculated based on alternative methodology." *Id*. Similarly, in

2021, the most recent year for which the full year data is available, "Commerce

applied an alternative comparison methodology for 15 companies (21 percent of

the total) and applied the A-to-A method for 58 companies, eight of which had a

zero rate." *Id*.

SeAH does not question the accuracy of these numbers, which demonstrate

that in almost 80 percent of determinations, the application of Cohen's *d* test did

not result in the application of the alternative comparison methodology. SeAH Br.

at 10, 26, 31-32. Instead, SeAH asserts that the "relevance of this analysis is

questionable" because the results found for other companies have no bearing on

results found for SeAH. SeAH Br. at 38. This may be true, but these numbers

refute SeAH's arguments that the Cohen's *d* test systematically produces

unreasonable results.

Instead, SeAH resorts to a straw man argument: SeAH contends that the

statute "does not direct Commerce to develop a methodology that would lead to a

change in the comparison methodology in a reasonable number of cases." SeAH

Br. at 39. SeAH misses the point. Commerce conducts DPA to "determine

whether the statutory criteria set forth in § 1677f-1(d)(1)(B) are met." *See e.g.,*

Appx14-15; Appx64-65 (explaining that the DPA is a "reasonable approach to

address the statutory requirements….").   In other words, Commerce is simply

implementing the statute and has developed the DPA to do so.

SeAH also contends that Commerce's statistics are "misleading" because

they are "skewed by the inclusion of investigations where the dumping margins

would have been above *de minimis* regardless of whether the A-to-A and A-to-T

methodology was used to calculate the margin." SeAH Br. at 41. Instead, SeAH

contends that Commerce should have excluded from its analysis all situations

where Commerce applied an alternative comparison methodology where the

resulting weighted-average dumping margins under both the average-to-average and alternative comparison methodologies were above *de minimis* and the rates differ by at least 25 percent. *Id.* at 41-42. SeAH provides a table purporting to analyze 59 instances of application of the DPA during the seven-year period between 2015-2021, contending that the DPA resulted in affirmative determinations in more than 50 percent of determinations. *Id.* at 42.

SeAH improperly limits its analysis to 59 respondents from the entire seven-year period between 2015 and 2021, whereas in 2021 alone Commerce applied its DPA to 73 respondents. Appx59. The mere fact that weighted-average dumping margins under both comparison methods are above *de minimis* does not end Commerce's DPA and Commerce could apply an alternative comparison methodology in such a scenario, if appropriate. "When the relative difference between the two rates is at least 25 percent, Commerce finds that the A-to-A method cannot account for such differences, and an alternative comparison methodology based on the A-to-T method may be warranted to calculate a company's weighted-average dumping margin." Appx90-91. Identically, when the weighted-average dumping margins cross the *de minimis* threshold, the A-to-A method cannot account for such differences. Under both definitions under the meaningful difference test, the A-to-A method cannot account for such differences and SeAH attempts to disregard how the test implements the statutory requirement.

SeAH attempts to distort the analysis by disregarding most investigations in which Commerce applied the differential pricing methodology. SeAH provides a table, in which the column "total investigations" is the sum of Commerce's final determinations in which the agency made negative findings and final determinations in which SeAH contends Commerce would have made a negative finding if it had applied the A-to-A method. SeAH Br. at 42. SeAH's table is misleading, because the "total investigations" column does not represent the total number of investigations during the relevant years. Out of the total 288 determinations in which Commerce applied the differential pricing methodology during the seven-year period, SeAH's "analysis" excluded 229 final determinations. SeAH Br. at 42.

Unlike SeAH, Commerce analyzed each instance in which it applied the DPA *to a respondent* without manipulating or limiting the data. Appx59. When the appropriate "total investigations"[6] value is used, the total percentage of final

---

[6] SeAH refers to "total investigations," but these numbers refer to the total number of respondents in the investigations. An investigation can involve multiple respondents, and, depending on the facts of each particular investigation, for some of these respondents the DPA could result in the application of the A-to-A methodology, while for others it could result in the application of an alternative comparison methodology.

determinations in which the application of the average-to-average method would have made the difference in the outcome is much lower than what SeAH presents.

Table 1: Application of an Alternative Comparison Methodology[7]

Year	Investigations: Number of Respondents (A)	*De Minimis* Rate Using the A-to-A Method (B)	Percent Use of an Alternative Comparison Methodology (C=B/A)
2015	18	3	16.67%
2016	52	5	9.62%
2017	35	9	25.71%
2018	48	1	2.08%
2019	28	2	7.14%
2020	34	5	14.71%
2021	73	6	8.22%
2015-2021 total:	**288**	**31**	**12.02%**

In other words, the respondents in which the application of the A-to-A method would have moved the weighted-average dumping margin below the *de minimis* threshold account for merely 12.02 percent of total final determinations during the seven-year period. Even for this relatively small number of respondents, Commerce's application of the alternative comparison methodology,

---

[7] (Source of data: SeAH Comments on Final Redetermination on Remand (June 14, 2022) (ECF 217), at Attachment 3.)

which unmasked dumping, was consistent with the statutory requirements, and

fulfills the congressional intent of the statute to unmask dumping.

The Cohen's *d* test does not look forward to the analysis in the meaningful

difference test, and the meaningful difference test does not look backwards to the

Cohen's *d* test to influence the results of either test. Accordingly, SeAH argument

that one should only consider situations where the weighted-average dumping

margin crosses the *de minimis* threshold because of alleged "false positives" in the

Cohen's *d* test is unfounded and illogical.

H.   The Court Should Disregard Materials And Arguments From Outside the Record of This Case

SeAH and amici disregard the well-established limitations on appellate

records by referencing non-record materials that were not part of the record that the

agency considered during the remand proceedings.[8] While amici did not

---

[8] The non-record materials cited by SeAH are:

- *Oxford Handbook of Quantitative Methods* (2013).
- Hedges and Olkin, *Overlap Between Treatment and Control Distributions as an Effect Size Measure in Experiments*, 21:1 PSYCHOLOGICAL METHODS (2016)
- Huberty and Lohman, *Group Overlap as a Basis for Effect Size*, 60:4 EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT (2000)
- J. Cohen, *A Power Primer*, 112:1 PSYCHOLOGICAL BULLETIN (1992)
- Ricca and Blaine, *Notes on a Nonparametric Estimate of Effect Size*, 90:1 JOURNAL OF EXPERIMENTAL EDUCATION (2022).

participate in the proceedings below, this Court's review is limited to the record

and arguments presented to the administrative agency and trial court, both by

statute and by the exhaustion and waiver/forfeiture doctrines, and this Court should

decline to consider amici's citations and arguments stemming from matters outside

the administrative record on appeal. 28 U.S.C. § 2637(d) (requiring that the Court

of International Trade "shall, where appropriate, require the exhaustion of

administrative remedies."); *see also Boomerang Tube*, 856 F.3d at 912. Courts take

"a 'strict view' of the requirement that parties exhaust their administrative

remedies . . . in trade cases." *Corus Staal BV v. United States*, 502 F.3d 1370,

1379 (Fed. Cir. 2007) ("{A}bsent a strong contrary reason, the court should insist

that parties exhaust their remedies before the pertinent administrative agencies.");

Boomerang Tube, 856 F.3d at 912 (same). "Simple fairness," moreover, "requires

---

The non-record materials cited by amicus curiae are:

- Edward L. Thorndike, *et al.*, *The Measurement of Intelligence* (1927)
- Larry V. Hedges, Ingram Olkin, *Overlap Between Treatment and Control Group Distributions as an Effect Size Measure in Experiments*, 21 Psychological Methods 61 (2016)
- Stephen Stigler, *The History of Statistics* (Harvard University Press 1986)
- Larry V. Hedges, *Review and Analysis of the Cohen's d Test as Used in the U.S. Department of Commerce's Differential Pricing Methodology* (Dec. 27, 2022)
- Figure 5

as a general rule that courts should not topple over administrative decisions unless

the administrative body not only has erred but has erred *against objection made at*

the time appropriate under its practice." *Mittal Steel*, 548 F.3d at 1383-84 (quoting

United States v. L.A. Tucker Truck Lines, Inc., 344 U.S. 33, 37 (1952) (emphasis

added by Court)). Thus, under the exhaustion doctrine, parties are required to raise

issues before Commerce at the time that Commerce is addressing them. *Id.*

(requiring exhaustion on remand). Amici should not be permitted, simply because

they have filed a brief as amici curiae to circumvent the standard of review

requiring that all records and argument be presented to the agency and raised for

the first time on appeal. *Axiom Res. Mgmt., Inc. v. United States*, 564 F.3d 1374,

1380 (Fed. Cir. 2009) ("The purpose of limiting review to the record actually

before the agency is to guard against courts using new evidence to 'convert the

"arbitrary and capricious" standard into effectively de novo review.'" (*quoting*

Murakami v. United States, 46 Fed. Cl. 731, 735 (2000))).

Likewise, the Court should decline to consider SeAH's citations and

arguments stemming from non-record documents. Appx8 n3; *see Camp v. Pitts*,

411 U.S. 138, 142 (1973) (explaining that "the focal point for judicial review

should be the administrative record already in existence, not some new record

made initially in the reviewing court"); *Axiom Res. Mgmt.*, 564 F.3d at 1380.

Limiting this Court's consideration to the record that was actually before the

agency is especially appropriate here because: first, SeAH has already attempted to rely on non-record materials and this Court affirmed Commerce's rejection of its earlier brief containing those non-record arguments; *see Stupp III*, 5 F.4th at 1348-51; and, second, on remand, Commerce allowed SeAH to add to the administrative record *all* non-record documents that formed the basis of SeAH's first appeal, which expanded the existing administrative record by more than 3,000 pages. Appx35. Allowing SeAH to add additional materials beyond those which it was already permitted is contrary to well-settled precedent. *Essar Steel Ltd. v. United States*, 678 F.3d 1268, 1277 (Fed. Cir. 2012) ("To allow constant reopening and supplementation of the record would lead to inefficiency and delay in finality."). Accordingly, the Court should either strike or disregard any references to or arguments based upon this non-record material. SeAH Br. at 18 n.25, 21 n. 31, 45 n.64 & 65, Appx6380-6395.

Finally, we note that SeAH included several arguments that were only made in footnotes. *See e.g.*, SeAH Br. at n. 25, 26, 38, 43, 47, 54, 78. For example, arguments regarding fluctuation of exchange rates, arguments regarding certain of SeAH's CONNUMS, arguments interpreting certain aspects of the academic literature, etc. These arguments have been waived. *SeAH Steel Vina Corp. v. United States*, 950 F.3d 833, 841 & n.5 (Fed. Cir. 2020) ("Arguments raised only in footnotes are waived.").

CONCLUSION

For these reasons, we respectfully request that the Court affirm the judgment

of the Court of International Trade.

<div align="right">

Respectfully submitted,

BRIAN M. BOYNTON
Principal Deputy Assistant Attorney General

PATRICIA M. MCCARTHY
Director

s/Claudia Burke
CLAUDIA BURKE
Deputy Director

s/Robert Ralph Kiepura
ROBERT RALPH KIEPURA
Trial Attorney
Commercial Litigation Branch
Civil Division
U.S. Department of Justice
P.O. Box 480
Ben Franklin Station
Washington, DC 20044
Tel.: (202) 305-4436
Fax: (202) 353-0461
Email: Robert.kiepura@usdoj.gov

</div>

November 6, 2023   Attorneys for Defendant

CERTIFICATE OF COMPLIANCE

Pursuant to Federal Rule of Appellate Procedure 32(g)(1), I certify under

penalty of perjury that the attached brief is proportionately spaced using Microsoft

Word, uses a Times New Roman typeface in 14 point font size, and, excluding the

parts of the brief exempted by Federal Rule of Appellate Procedure 32(f), contains

13,985 words using the word count function of the word processing software used

to prepare this brief (Microsoft Word).

/s/ Robert R. Kiepura
ROBERT R. KIEPURA
Trial Attorney
Commercial Litigation Branch
Department of Justice, Civil Division
P.O. Box 480
Washington, DC 20044
Telephone: (202) 305-4436
Robert.Kiepura@usdoj.gov
November 6, 2023   Attorney for the Defendant-Appellee

FORM 30. Certificate of Service

<div align="right">Form 30
July 2020</div>

UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT

<u>CERTIFICATE OF SERVICE</u>

Case Number <u>23-1663</u>

Short Case Caption <u>Stupp v. U.S.</u>

NOTE: Proof of service is only required when the rules specify that service must be accomplished outside the court's electronic filing system. See Fed. R. App. P. 25(d); Fed. Cir. R. 25(e). Attach additional pages as needed.

I certify that I served a copy of the foregoing filing on <u>11/06/2023</u>

by ☐ U.S. Mail   ☐ Hand Delivery   ☑ Email   ☐ Facsimile
    ☐ Other: _____

on the below individuals at the following locations.

Person Served	Service Location (Address, Facsimile, Email)
Jeffrey Winton	jwinton@winton.law
Jooyoun Jeong	jjeong@winton.law
Jeff Gerrish	jgerrish@scharginassociates.com
Roger Schargin	rschagrin@schagrinassociates.com

☑ Additional pages attached.

Date: <u>11/06/2023</u>

Signature: <u>/s/ Robert R. Kiepura</u>

Name: <u>Robert R. Kiepura</u>

CERTIFICATE OF SERVICE

Pursuant to Federal Rule of Appellate Procedure 25(d), I certify under penalty of perjury that on this 6th day of November, 2023 a copy of the Nonconfidential version of the foregoing "Brief of Defendant-Appellee" was filed electronically. I understand that notice of this filing will be sent to all parties by operation of the Court's electronic filing system. Parties may access this filing through the Court's system.

On November 1, 2023, Jooyoun Jeong on behalf of SeAH and Jeff Gerrish on behalf of Welspun, stated via email that they consented to service of the confidential version of the forgoing brief via email. Accordingly, a copy of the confidential version of this brief has been served on this day, November 6, 2023, via email upon Jeffrey Winton and Jooyoun Jeong for SeAH and Roger Schagrin and Jeff Gerrish for Welspun.

None of the other parties listed in the caption of this case are covered by either the administrative protection order or judicial protective order applicable to this case. Accordingly, a copy of the forgoing brief was not served upon any other party.

/s/ Robert R. Kiepura

ROBERT R. KIEPURA
Trial Attorney
Commercial Litigation Branch
Department of Justice, Civil Division
P.O. Box 480

Washington, DC 20044
Telephone: (202) 305-4436
Robert.Kiepura@usdoj.gov

November 6, 2023   Attorney for the Defendant-Appellee

UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT

CERTIFICATE OF CONFIDENTIAL MATERIAL

Case Number: 23-1663

Short Case Caption: Stupp v. U.S.

> **Instructions:** When computing a confidential word count, Fed. Cir. R. 25.1(d)(1)(C) applies the following exclusions:
>
> - Only count each unique word or number once (repeated uses of the same word do not count more than once).
>
> - For a responsive filing, do not count words marked confidential for the first time in the preceding filing.
>
> The limitations of Fed. Cir. R. 25.1(d)(1) do not apply to appendices; attachments; exhibits; and addenda. *See* Fed. Cir. R. 25.1(d)(1)(D).

The foregoing document contains __18__ number of unique words (including numbers) marked confidential.

☐ This number does not exceed the maximum of 15 words permitted by Fed. Cir. R. 25.1(d)(1)(A).

☑ This number does not exceed the maximum of 50 words permitted by Fed. Cir. R. 25.1(d)(1)(B) for cases under 19 U.S.C. § 1516a or 28 U.S.C. § 1491(b).

☐ This number exceeds the maximum permitted by Federal Circuit Rule 25.1(d)(1), and the filing is accompanied by a motion to waive the confidentiality requirements.

Date: 11/06/2023

Signature: /s/ Robert R. Kiepura

Name: Robert R. Kiepura